## **CLAIMS**

## What is claimed is:

SUB AB>

5

6

A file content classification system comprising:

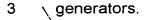
a digital ID generator;

an ID appearance database coupled to receive IDs from the ID generator; and

a characteristic comparison routine identifying the file as having a characteristic based on ID appearance in the appearance database.

- The content classification system of claim 1 wherein said ID
   generator comprises a hashing algorithm.
- 1 3. The content classification system of claim 2 wherein said hashing algorithm is the MD5 hashing algorithm.
- The content classification system of claim 1 wherein said ID
   appearance database tracks the frequency of appearance of a digital ID.
- The content classification system of claim 1 further including a plurality of digital ID generators on different systems all coupled to and providing IDs to said ID appearance database.
- 1 6. The content classification system of claim 5 wherein said plurality of digital ID generators are coupled to said database via a combination of public and private networks.
- 7. The content classification system of claim 6 wherein said database is coupled to an intermediate server which is coupled to said plurality of





1 2

3

4

5

6

- 1 8. The content classification system of claim 6 wherein said intermediate server is a web server.
- 9. The content classification system of claim 1 wherein said characteristic comprises junk e-mail and said characteristic is defined by a frequency of appearance of a digital ID.
  - 10. A method for identifying a characteristic of a data file, comprising:
    generating a digital identifier for the data file and forwarding the identifier to a processing system;

determining whether the forwarded identifier matches a characteristic of other identifiers; and

- processing the email based on said step of determining.
- 1 11. The method of claim 10 wherein said step of generating comprises 2 hashing at least a portion of the data file.
- 1 12. The method of claim 11 wherein said step of hashing comprises 2 using the MD5 hash.
- 1 13. The method of claim 11 wherein said step of generating comprises 2 hashing multiple portions of the data file.
- 14. The method of claim 10 wherein said data file is an email message
   2 and said step of determining comprises determining whether said email is
   3 spam.
- 1 15. The method of claim 10 wherein said step of determining identifies

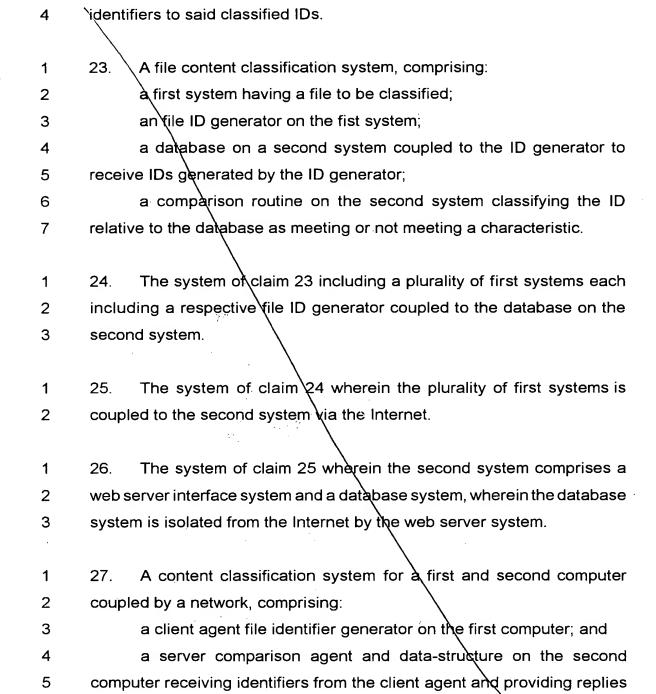


- said e-mail as spam by tracking the rate per unit time a digital ID is generated.
- 1 16. The method of claim 10 wherein said step of generating comprises
- 2 generating Ds at a plurality of source systems all coupled via a network
- 3 to at least one processing system performing the determining step.
- 1 17. The method of claim 16 wherein said step of processing comprises
- 2 instructing said plurality of source systems to perform an action with the
- 3 email based on said determining step.
- 1 18. A method of filtering an email message, comprising:
- 2 processing the message to provide a digital identifier;
- 3 comparing the digital identifier to a characteristic database of digital
- 4 identifiers to determine whether the message has said characteristic; and
- 5 processing the message based on said step of comparing.
- 1 19. The method of claim 18 wherein said step of processing occurs on
- 2 at least one first system, and said step of comparing occurs on a second
- 3 system.
- 1 20. The method of claim 19 wherein said step of processing occurs on
- 2 a plurality of first systems.
- 1 21. The method of claim 19 wherein said at least one first system and
- 2 second system are coupled by the Internet.
- 1 22. The method of claim 18 wherein said step of comparing comprises
- determining the frequency of a particular ID occurring in a time period,
- 3 classifying said ID as having a characteristic, and comparing digital

6

7

8



wherein the client agent processes the file based on replies from the

server comparison agent.

to the client agent;





1	28.	A method for providing a service on the Internet, comprising:
2		collecting data from a plurality of systems having a client agent on
3	the In	ternet to a server having a database;
4		characterizing the data received relative to information collected in
5	the da	atabase; and
6		transmitting a content identifier to the client agent.
1	29.	The method of claim 28 wherein said step of collecting comprises
2	collec	ting a digital identifier for a data file.
1	30.	The method of claim 28 wherein said data file is an e-mail.
1	31.	The method of claim 29 wherein said step of characterizing
2	comp	rises:
3		tracking the frequency of the collection of a particular identifier;
4		characterizing the data file based on said frequency;
5		storing the characterization; and
6		comparing collected identifiers to the known characterization.